CrossMark

ORIGINAL ARTICLE

Prediction of the hot metal silicon content in blast furnace based on extreme learning machine

Haigang Zhang[1,2] · Sen Zhang[1,2] · Yixin Yin[1,2] · Xianzhong Chen[1,2]

Abstract Silicon content in hot metal is an important indicator for the thermal condition inside the blast furnace in the iron-making process. The operators often refer the silicon content and its change trend for the guidance of next production. In this paper, we establish the neural network model for the prediction of silicon content in hot metal based on extreme learning machine (ELM) algorithm. Considering the imbalanced operating data, weighted ELM (W-ELM) algorithm is employed to make prediction for the change trend of silicon content. The outliers hidden in the real production data often tend to undermine the accuracy of prediction model. First, an outlier detection method based on W-ELM model is proposed from a statistical view. Then we modified the ordinary ELM and W-ELM algorithms in order to reduce the interference of outliers, and proposed two enhanced ELM frameworks respectively for regression and classification applications. In the simulation part, the real operating data is employed to verify the better performance of the proposed algorithm.

✉ Sen Zhang
zhangsen@ustb.edu.cn

1 School of Automation and Electrical Engineering, University of Science and Technology Beijing(USTB), No. 30, Xueyuan road, Haidian District, Beijing 100083, China

2 Key Laboratory of Advanced Control of Iron and Steel Process (Ministry of Education), University of Science and Technology Beijing, Beijing 100083, China

1 Introduction

The iron and steel production is an important pillar industry of national economy in China, including mining, beneficiation, sintering, iron-making and steel-making et al. [1]. The blast furnace (BF) iron-making process is the most significant step in the manufacture of iron and steel. The BF is a huge reaction vessel, which converts iron oxides into liquid iron (hot metal) through many chemical and physical reactions [2]. The iron-making process pursues stable operation. Anterograde blast furnace can not only prolong life, but also raise productivity. It is a prerequisite to keep proper temperature in BF for the stable operation of iron-making process. The over-cold and over-hot situation in BF will affect the production process, and even lead to production failures [3, 4].

The density of silicon in hot metal has always been an important instruction of the blast furnace operation. The silicon content is not only an important parameter and basis for the heat in BF, but it also is a direct response to the status and trend of hot metal [5–7]. In iron-making production, the operators often adjust the temperature in BF based on the change amplitude, trends and frequency of silicon content, resulting of the reduce of the volatility of furnace conditions. In the smelting process, if we can grasp the silicon contents as well as its trend in a timely manner, it can well improve the quality of pig iron and reduce the coke rate through the corresponding operation.

Taking into account of the complex 'black box' character in the measurement of silicon content, the mechanism modeling becomes impractical. In such case, the data-driven technique is a good choice. There are many methods for the prediction of silicon content based on data-driven methods. We summarize that the prediction model of silicon content can be divided into linear-time series models

🖄 Springer

1698

Int. J. Mach. Learn. & Cyber. (2018) 9:1697–1706

and non-linear-time series models. For the linear models, a state-space model was developed for the prediction of hot metal silicon content, using the blast furnace pressure and the over-all heat loss as input variables [8]. Bhattacharya [3] established the hot metal silicon prediction model using partial least squares (PLS) algorithm. Other linear prediction model can be referred to [10–12]. There are many non-linear time-series models employed in the prediction of hot metal silicon content [5] applied a pruning algorithm to find relevant inputs and an approximate network connectivity based on feedforward neural network, and then made prediction for the silicon content in BF. Support vector machine (SVM) and its variant least square support vector machine (LS-SVM) are often employed in not only the value prediction, but also the change trend classification of the silicon content in BF [1, 6, 13]. Tang applied the chaos particle swarm optimization to select the optimal parameters of support vector regression (SVR) for the prediction of hot metal silicon content [14]. Other nonlinear time-series models, such as wiener models [15] and chaos models [16], are also applied in the prediction of hot metal silicon content.

Extreme learning machine (ELM) is a novel single hidden layer neural networks algorithm [9, 17]. It has drawn more and more attention because of the fast training speed and good generalization performance. The essence of ELM algorithm is the random selection of model learning parameters, while the output weights can be calculated by the ordinary least square (LS) method despite of some iterative ways [18, 19]. Huang has proved the universal approximation capability and classification capability of ELM model, which makes a support for the improvement of ELM algorithm [20]. Since proposed, ELM algorithm has experienced a period of rapid development, such as from batch learning to online sequential learning, from sigmoid and RBF activation functions to almost any nonlinear piecewise activation function, from single hidden layer feedforward networks to multi hidden layers feedforward networks etc [21, 22]. Nowadays, ELM algorithm has been widely employed in many real applications such as regression, classification, clustering, feature learning and so on [23].

Raw data with imbalanced class distribution can be found almost everywhere, especially in industrial applications [24]. In iron-making process, stable production is the common goal of the operator, in which case, BF is often in an anterograde state. In most of the production condition, the dramatic change of silicon content often seldom occur. Thus the obtained data contains a large amount of unbalanced classes. Weighted ELM algorithm (W-ELM) can well deal with data with unbalanced class distribution, which is suitable for the silicon content trend prediction [20]. In addition, outliers occur frequently in the complex industrial process and can have serious unsatisfied consequences [30] presents

an outlier-robust extreme learning machine algorithm, where the ℓ_1-norm loss function is used to enhance the robustness. However, it is very suitable to the regression applications. In this paper, we modify ELM and W-ELM algorithms and propose two schemes to deal with outliers. One is the outlier detection method based on W-ELM model, where a statistic method is employed to detect the outliers in real BF data. Then the enhanced ELM and W-ELM algorithms (named E-ELM and EW-ELM) are proposed respectively for regression and classification applications in the prediction model of silicon content in hot metal.

The rest of this paper is organized as follows. Section 2 presents a review of the related works including the ordinary ELM and W-ELM algorithms. In Sect. 3, we present the outlier detection method, and propose two robust ELM frameworks for regression and classification applications. The prediction model for silicon content is presented in Sect. 4. Simulation results will be shown in Sect. 5, and Sect. 6 is the conclusion of our work.

2 Brief Review of ELM and W-ELM

2.1 ELM

The ELM algorithm was originally proposed by Huang and it can be regarded as a general single hidden layer network (SLFNs) [17]. ELM gets rid of human tuning with random initialization of SLFNs learning parameters. Then the output weights can be determined by the theory of least square method [18].

Given a training set consisting of N arbitrary distinct samples $S = \left\{ (x_i, t_i) | x_i \in R^n, t_i \in R^m, i = 1, 2, \ldots, N \right\}$, where x_i and t_i represent the features of model input and output respectively, the SLFNs network function with $\tilde{N}$ hidden nodes can be formulated as

$$f_{\tilde{N}} = \sum_{j=1}^{\tilde{N}} \beta_j G(a_j, b_j, x_i) = t_i, \quad i = 1, 2, \ldots, N \qquad (1)$$

where a_j and b_j are the learning parameters which will be determined randomly. β_j is the output weight matrix connecting the jth hidden and the output nodes. $G(a_j, b_j, x_i)$ is a nonlinear piecewise continuous function which satisfies ELM universal approximation capability.

The above $\tilde{N}$ equations can be written in matrix form as

$$H\beta + \varepsilon = T \qquad (2)$$

where $H = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_{\tilde{N}}, b_{\tilde{N}}, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_{\tilde{N}}, b_{\tilde{N}}, x_N) \end{bmatrix}_{N \times \tilde{N}}$ is called

output hidden layer matrix. ε represents the white noise, $\varepsilon \in N(0, \sigma^2)$.

Int. J. Mach. Learn. & Cyber. (2018) 9:1697–1706

1699

ELM aims to minimize the training error as well as the norm of the output weights. Then ELM model with a single-output node can be formulated as:

$$minimize \quad J_{ELM} = \frac{1}{2}\|\beta\|^2 + C\frac{1}{2}\sum_{i=1}^{N}\|e_i\|^2$$
$$subject\ to: \quad h(x_i)\beta = t_i - e_i \tag{3}$$

where e_i represents the training error of the ith observation. $h(x_i)$ is the ith row of output hidden layer matrix. C is a use-specified parameter and it can provide a tradeoff between the training error and the norm of the output weights.

The output weights of ELM algorithm can be calculated by LS method. In the case where the number of training observations is larger than that of the hidden nodes, it means that H is of full column rank. Then the output weight should be estimated as follows:

$$\hat{\beta} = \left(\frac{I_{\tilde{N}}}{C} + H^T H\right)^{-1} H^T T \tag{4}$$

where $I_{\tilde{N}}$ is an identity matrix with dimension $\tilde{N}$ (similarly hereinafter). When the number of training patterns is smaller than the hidden nodes' number, H has more columns than rows. H has the full row rank and HH^T is nonsingular. Then we have

$$\hat{\beta} = H^T\left(\frac{I_N}{C} + HH^T\right)^{-1} T \tag{5}$$

Notably, $\frac{I_{\tilde{N}}}{C}$ or $\frac{I_N}{C}$ is called the regularization parameter, which sets up the connection between ELM with other statistical theories like matrix theory and ridge regression [20].

2.2 W-ELM

The unbalanced datasets often appear in many real world applications, especially in many industrial areas with the pursuit of "stable production". The unbalanced distribution of training and testing datasets would deteriorate the ELM performance, which motivates the emergence of W-ELM algorithm [20]. Here we modify the performance index of ELM model as

$$minimize \quad J_{ELM} = \frac{1}{2}\|\beta\|^2 + CW\frac{1}{2}\sum_{i=1}^{N}\|e_i\|^2$$
$$subject\ to: \quad h(x_i)\beta = t_i - e_i \tag{6}$$

where W is an $N \times N$ diagonal weighted matrix associated with every training sample. The ith element w_i on the diagonal represents the weight allocated to the ith sample x_i. Generally, one can put relatively larger weights on the minority classes, and vice versa.

Modeled on the above derivation, one can obtain the following solutions:

$$\begin{cases} \hat{\beta} = \left(\frac{I_{\tilde{N}}}{C} + H^T WH\right)^{-1} H^T WT, \ \tilde{N} < N \\ \hat{\beta} = H^T\left(\frac{I_N}{C} + WHH^T\right)^{-1} WT, \ \tilde{N} > N \end{cases} \tag{7}$$

W-ELM imports the weighted matrix W into the performance index and establishes a balance between the majority and minority sample classes. The suitable weights pushed toward different samples are necessary for the generalization performance. In general, each weight is inversely proportional to the error variance, which can reflect the information in that observation. An observation with small error variance has a large weight since it contains relatively more information than an observation with large error variance (small weight). In the classification applications, W is often determined based on the number of observations belonging to the same class. So

$$w_i = 1/\#n_i \tag{8}$$

where $\#n_i$ represents the number of samples belong to the i th class. The majority classes with relatively large number of observations have relatively small weights, while large weights are assigned on the minority classes. Such weight determination can well exclude the imbalanced phenomenon among the samples.

3 Outlier detection and robust ELM framework

In this section, we present two schemes to deal with the outliers. One is focused on the outlier detection from a statistical point. Then we propose an enhanced (weighted) extreme learning machine algorithm (E-ELM and EW-ELM). E-ELM can be well applied in the value prediction of silicon content in hot metal, while EW-ELM is a good choice to make the change trend classification of silicon content.

3.1 Outlier detection

An outlier is an observation point that appears to deviate markedly from other observations. Compared with the normal samples, the outlier often presents some non-interpretable and non-coordination features. Visual inspection is a generally helpful tool for outlier detection from a global perspective. However, considering the large and complex datasets, the automated outlier detection procedures are necessary [26].

In this subsection, we present a statistical scheme to detect the outliers subject to W-ELM algorithm [20]. In

1700

Int. J. Mach. Learn. & Cyber. (2018) 9:1697–1706

most complex industrial applications, data tends to be contaminated. The W-ELM model introduced in the above section can be modified as:

$$T = H\beta + \varepsilon_w \tag{9}$$

where $\varepsilon_w \in N(0, \sigma^2/W)$ and W is the weighted matrix defined above.

Then the output estimation value can be calculated as following:

$$\hat{T}_w = H\hat{\beta}_w = H(H^T W H)^{-1} H^T W T \overset{\Delta}{=} Q_w W T \tag{10}$$

where $Q_w = H(H^T W H)^{-1} H^T$, and we omit the regularization factor C for simplicity. It is easy to get that Q_w is symmetric ($Q_w^T = Q_w$) and idempotent ($Q_w W Q_w = Q_w$).

Next one can define the estimated error as $e = T - \hat{T}_w = (W^{-1} - Q_w)WT$. So the corresponding variances can be calculated as:

$$\begin{cases} V(\hat{\beta}_w) = \sigma^2 (H^T W H)^{-1} \\ V(\hat{T}_w) = V(Q_w W T) = \sigma^2 Q_w \\ V(e) = \sigma^2 (W^{-1} - Q_w) \end{cases} \tag{11}$$

Since

$$E(e^T W e) = E[tr(e^T W e)] = \sigma^2 \cdot tr\{W(W^{-1} - Q_w)\} = \sigma^2(N - \tilde{N}) \tag{12}$$

One can get the unbiased estimation of σ^2:

$$\hat{\sigma}^2 = \frac{e^T W e}{N - \tilde{N}} \tag{13}$$

In the outlier detection method, the internally and externally studentized weighted residuals are taken into consideration. A relative lemma is presented as following:

Lemma 1 *Considering the data sequential without the ith observation, $\hat{\sigma}_i^2$ denotes the estimation of square deviation for the remaining samples, which can be calculated as*

$$\hat{\sigma}_{(i)}^2 = \frac{(N - \tilde{N})\hat{\sigma}^2 - w_i e_i^2 / (1 - w_i q_{w,i})}{N - \tilde{N} - 1} \tag{14}$$

where $e_i = t_i - h_i^T \hat{\beta}_w$ and $q_{w,i} = h_i^T (H^T W H)^{-1} h_i$, where h_i^T is the ith row vector of hidden layer matrix H. And we present the main proof in Part A in the Appendix.

Then for the outlier detection procedure, we present the following theorem:

Theorem 1 *For a data sequence with normal distribution, the outlier can be detected with relative large internally and externally studentized weighted residuals*

(denoted by $S_{w,i}$ and $S'_{w,i}$). And the $S_{w,i}$ and $S'_{w,i}$ for the ith observation can be obtained as

$$\begin{aligned} S_{w,i} &= \frac{\sqrt{w_i} e_i}{\hat{\sigma}_i \sqrt{1 - w_i q_{w,i}}} \\ S'_{w,i} &= \frac{\sqrt{w_i} e_i}{\hat{\sigma} \sqrt{1 - w_i q_{w,i}}} \end{aligned} \tag{15}$$

Note: In statistics, a studentized residual is the quotient resulting from the division of a residual by an estimate of its standard deviation, which is an important technique in the detection of outliers. The ordinary calculation of studentized residual is $S_i = e_i / (\hat{\sigma}\sqrt{1 - q_i})$. In Theorem 1, we added the corresponding weight to every observation. Then the calculations of studentized weighted residuals are modified as (15). The basic idea is to delete the observations one at a time, each time fitting the regression model on the reminding $N - 1$ observations. Then we compare the observed response values to their fitted values based on the models with the ith observation deleted. The outliers would present the characteristic of non-random draft. The observations with relative large studentized residuals would be regarded as the outliers. More detail information about this theorem can be referred in [27, 31].

3.2 E-ELM and EW-ELM

In the above subsection, we present the outlier detection schemes based on the statistics analysis. It is very essential and useful to mark the outliers in the data preprocessing stage. Now we improve the ordinary ELM and W-ELM algorithms in order to make a double guarantee for effect attenuation of the outlier.

In the data mining and modeling process, it is necessary to define a cost function or performance index to measure the performance of the trained model [32]. As shown in (weighted) ELM theory, the output weights are calculated analytically based on the minimum of least mean squares (LMS) of the residuals. However, data modeling through the LMS method is often affected by the abnormal points, providing the incorrect results. Here we propose a robust error measure for ELM and W-ELM algorithm with outliers through the modification of cost function and performance index.

For brief description, we present the abbreviated form of the cost function in ELM algorithm by omitting the minimum term of $\|\beta\|^2$ and revising the squared error to mean value. That is

$$\varphi = \frac{1}{2N} \sum_{i=1}^{N} \|e_i\|^2 = \frac{1}{N} \sum_{i=1}^{N} \xi_i \tag{16}$$

Int. J. Mach. Learn. & Cyber. (2018) 9:1697–1706

1701

where $\xi_i = \frac{1}{2} \|e_i\|^2$ called the error function, which is sensitive to the training error of outliers. Next we present the modified least mean log squares(LMLS) method subject to ELM cost function, and the modified cost function and error function are as follows [24]:

$$
\begin{cases}
\psi = \frac{1}{N} \sum_{i=1}^{N} \log\left(1 + \frac{1}{2}\|e_i\|^2\right) = \frac{1}{N} \sum_{i=1}^{N} \xi_i \\
\xi_i = \log\left(1 + \frac{1}{2}\|e_i\|^2\right)
\end{cases}
\tag{17}
$$

Then the following theorem for the proposed E-ELM and EW-ELM algorithms is presented subject to interference elimination of outliers.

Theorem 2 *Considering the ordinary ELM algorithm, the cost function with LMLS method is more robust than with LMS subject to the gross errors for the outliers. In other words, LMLS method presents less sensitive to the residuals of outliers than the ordinary LMS method.*

Proof In most neural network structure including ELM model, we aim to search for the optimal output weights. We define $\vec{v}$ as the modifiable network parameter vector and the directional differentiating with respect to $\vec{v}$ for the LMS method yields:

$$
\frac{\partial \varphi}{\vec{v}} = \frac{1}{N} \sum_{i=1}^{N} \tau(e_i) \frac{\partial e_i}{\vec{v}}
\tag{18}
$$

where $\tau(e_i) = \frac{\partial \xi_i}{\partial e_i} = e_i$, which tends to zero with the iterative optimization. One can call $\tau(e_i)$ the influence function because it reflects the influence degree of the ith individual. Similarly, one can calculate the influence function for the LMS and LMLS methods as follows:

$$
\begin{cases}
\tau_{ELM}(e_i) = \frac{\partial \xi_{ELM-i}}{\partial e_i} = e_i \\
\tau_{E-ELM}(e_i) = \frac{\partial \xi_{E-ELM-i}}{\partial e_i} = \frac{e_i}{1 + \frac{1}{2}\|e_i\|^2}
\end{cases}
\tag{19}
$$

Figure 1 presents the influence function of LMS and LMLS methods subject to different model errors. Obviously, the influence function of LMS method changes monotonically, which is more sensitive to the model error. However, the magnitude of LMLS influence function reaches its maximum and starts to decrease for relatively large residuals. From the extreme perspective, the outliers have no effect to the ELM model with LMLS influence function due to.

$$
\lim_{e_i \to \infty} \tau_{E-ELM}(e_i) = \lim_{e_i \to \infty} \frac{e_i}{1 + \frac{1}{2}\|e_i\|^2} = 0
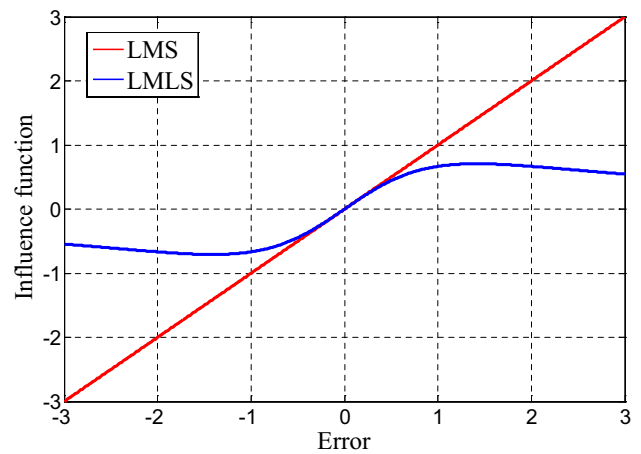\tag{20}
$$

Fig. 1 The comparison of influence function between LMS and LMLS method

The proposed E-ELM and EW-ELM algorithms applying the cost function with LMLS method can respectively deal with the regression and classification applications with outliers. One can employ the ordinary gradient decent method to train the proposed ELM frameworks. In part B of the "Appendix", the E-ELM and EW-ELM algorithms, compared with the ordinary ELM algorithm and the OR-ELM algorithm proposed in [30], present more robust and satisfied performance. However, the proposed algorithms take the relatively long training time. Thus E-ELM and EW-ELM algorithms can be well applied in the applications with serious outliers and no strict requirement for the training time.

4 The silicon content forecast model

The iron-making process in BF strives for stable production. The silicon content in hot metal can well reflect the production status in the interior of BF. Here we propose the prediction scheme for the silicon content in hot metal based on the real production data.

The influence parameters selection of the silicon content prediction model is an important link. Suitable feature parameters are necessary to improve the prediction accuracy [29]. There are two categories of blast furnace parameters influencing the silicon content. One is the state parameters, including blast volume, feeding speed et al., while the other one is the control parameters, such as blast temperature, blast pressure et al. Here, we carry out the correlation analysis among the silicon content with other blast furnace parameters, and choose the input parameters with relatively high correlation with the silicon content. Based on the operating experience, we chose 9 parameters as shown in Table 1.

Table 1 The inputs of prediction model

Variable name	Unit
Blast volume	m^3/min
Blast temperature	°C
Blast pressure	kPa
Pulverized coal injection	ton
Oxygen enrichment percentage	wt%
Feed speed	mm/h
Smelting intensity	$t/m^3 \cdot d$
The latest silicon content	et%
Gas permeability	$m^3/\text{min} \cdot kPa$

As described above, the trend of change of silicon content is also very important for the blast furnace operation. Based on the operation experience, we classify the silicon content change range into five categories as shown in Table 2. The range boundaries of silicon content variation are delineated based on the operators' experience. The last row in Table 2 shows the number of real data in each category of silicon content. Apparently, most of the silicon contents are in slight changes, while the sharp descent and assent situations rarely occur. ELM algorithm can not obtain the satisfied prediction results subject to such imbalanced data.

In the prediction model of silicon content in BF, we employ the outlier detection method described in Theorem 1 to identify the abnormal data. The proposed E-ELM and EW-ELM algorithms are applied to make prediction for the value and change trends of silicon content.

5 Simulation results

In this section, we present the simulation results based on the real iron-making data. All the simulations have been conducted in Matlab 7.8.0(2009a) running on a desktop PC with AMD Athlon(tm) II X2 250 processor, 3.00-GHz CPU and 2G RAM. All the simulation results are averaged 50 times.

5.1 Data preprocessing

The real production data collected in a blast furnace with 2500 m³ are employed in the experiment. We choose 1205 sets of silicon content data. Because of the complex operation environment in BF production, the sampled data tent to be contaminated by various forms of noise, such as process noise and measurement noise. Denoising operation is necessary in data preprocessing. In addition, there exist big delays in iron-making process in BF. The uncoordinated input variables tend to destroy the accuracy of prediction model. In the data preprocessing, we carry out the correlation analysis between the silicon content and inputs of prediction model at different time, and determine the delay time under the condition of maximum correlation. The value of silicon content in hot metal and the inputs of prediction model are not in the same order of magnitude. Some indicators have big difference. For instance, the blast temperature is about 1000 °C, while the change of silicon content is very slight. It is necessary to make standardization for data procession.

5.2 Parameter configuration

ELM algorithm can apply any nonlinear piecewise continuous function which satisfies ELM universal approximation capability. In the simulation, we employ the sigmoidal additive activation function $G(a, b, x) = 1/(1 + \exp(-(a \cdot x + b)))$, where the input weights and biases are randomly generated from the range $[-1, 1]$. Based on the theory of ELM algorithm, the model becomes more and more accurate with the increase of the number of hidden nodes. Taking into account the big delay phenomenon of iron-making process and the relatively long sampling period, we do not care about the training time of ELM prediction model. Figure 2 presents the ELM model prediction errors with different number of hidden nodes. When the number of hidden nodes is less than 200, the prediction error is relatively large, and not stable. Then with the increase of hidden nodes, the prediction error presents a downward trend, and the trend gradually flattens out.
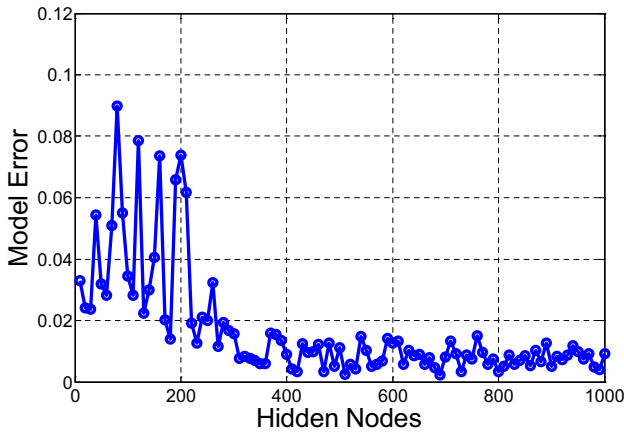
Table 2 The change trend class of silicon content

Class	Sharp descent	Slight descent	Steady	Slight assent	Sharp assent
Silicon change range	$\Delta[Si] \leqslant -0.2$	$-0.2 < \Delta[Si] < -0.05$	$-0.05 < \Delta[Si] < 0.05$	$0.05 \leqslant \Delta[Si] < 0.2$	$\Delta[Si] \geqslant 0.2$
No.	132	318	292	335	128

Fig. 2 The model error with different number of hidden nodes

Fig. 4 The internally and externally studentized weighted residuals

Fig. 3 The silicon content data series with outliers

Fig. 5 Comparison results for the regression of silicon content
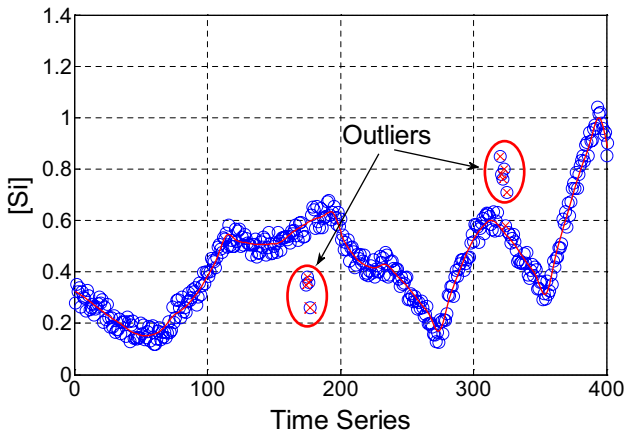
5.3 Simulation results

1. *Outlier detection*: The outliers may exist in the obtained data, which can deteriorate the performance of ELM model. Here we present the simulation results of the outlier detection method. The statistical method is employed to make outlier detection for the real silicon content data. Figure 3 presents partial silicon content data, where the data points encircled by red crosses are the outliers. Visually, the outliers do not meet the overall change trend of the real silicon content data. Figure 4 shows the internally (the blue line) and externally (the green line) studentized weighted residuals. The outliers have relatively large internally and externally studentized weighted residuals, marked by red circles.

2. *Silicon content regression*: In this simulation, the proposed E-ELM algorithm is employed in the value prediction of silicon content. In order to verify the better
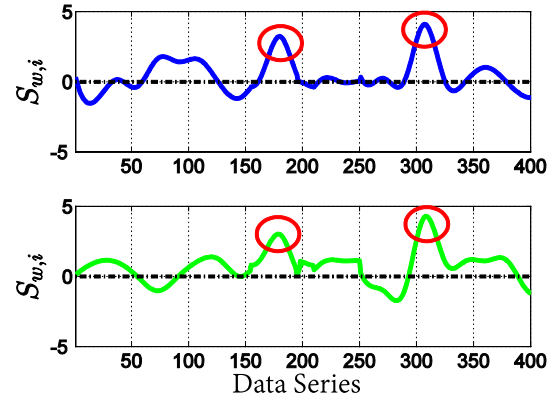
generalization performance of E-ELM algorithm, we compared it with the ordinary ELM algorithm and LS-SVM algorithm, which is often employed in the prediction of silicon content such as in [1, 6]. For LS-SVM algorithm, the famous Gaussian kernel function is applied: $k(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right)$, where γ is the kernel parameter. In ELM and E-ELM prediction model, the number of hidden nodes is set as 800. Figure 5 presents the simulation results. The blue line shows the change of silicon content in hot metal, where one can see it has relatively large fluctuations. The output of E-ELM algorithm (represented by the red line) can well track the change of silicon content, while the performance of ordinary ELM algorithm is not satisfactory. In addition, LS-SVM algorithm obtains the worst performance represents by the green line.

1704

Int. J. Mach. Learn. & Cyber. (2018) 9:1697–1706

Table 3 The simulation results of silicon content prediction

Algorithm	Parameter		Regression (RMSE)	Classification hit rate (%)
SVM/LS-SVM	$\gamma = 2$		0.0256	89.56
ELM	#Nodes	500	0.0167	93.45
		800	0.0156	93.23
		1000	0.0152	94.58
E-ELM		500	0.0142	95.23
EW-ELM		800	0.0101	96.12
		1000	0.0100	96.15

Table 3 shows the mean square error (MSE) in this simulation, where E-ELM algorithm obtains the least MSE. Taking into account the special environment in iron-making production, the training time of E-ELM algorithm can satisfied the demand of iron-making operations. From the above simulation, there are some outliers in the training data. E-ELM can well get rid of the interference of outliers, and obtains the best performance in the value prediction of silicon content in hot metal.

3. *Silicon content classification*: In this part, we present the simulation results of the classification for the change trends of silicon content. The proposed EW-ELM algorithm is employed, while the weighted matrix *W* is selected based on Eq. (8). In order to verify the better classification performance of EW-ELM algorithm, we compared EW-ELM algorithm with SVM algorithm. The same as above, kernel function is employed in SVM algorithm. SVM is an excellent binary classification algorithm, while one should take approximate measures in data preprocessing subject to multi classification applications. Here the one-against-all method is applied to make multi-classification for the change trends of silicon content. The simulation results are evaluated by the following hit rate criteria:

$$\text{Hit rate} = \frac{Count(\text{correct classifications on the testing set})}{Count(\text{testing set})}$$

Table 3 presents the comparison results for the classification of silicon content based on SVM, ELM and EW-ELM algorithms. The EW-ELM algorithm obtains the best hit rate. EW-ELM can well exclude the effect of outliers to the model. In summary, EW-ELM can obtain the best classification performance for the change trends of silicon content, and well satisfy the need of industrial production.

6 Conclusion

The prediction of silicon content in hot metal is important for the quality of the iron, but also an indicator of the operating condition of BF production. The operators often carry out the operations based on the value and change trends of silicon content. In this paper, we proposed a novel prediction scheme for the silicon content of hot metal based on ELM algorithm. Because of the complex iron-making environment, there are many outliers exist in the obtained data, which can deteriorate the performance of ELM algorithm. We proposed an outlier detection method from a statistical point of view. Then two modified ELM algorithms are proposed to get rid of the interference of outliers, named E-ELM and EW-ELM. It is worth noting that the proposed ELM frameworks can be applied to any regression and classification application with outliers in machine learning fields. The E-ELM algorithm is applied in the value prediction of silicon content, while EW-ELM algorithm makes prediction for the change trend of silicon content. In the simulation part, the real BF production data has been employed to verify the satisfied performance of the proposed scheme to make prediction of silicon content.

Appendix

Part A: Proof for Lemma 1

In this part, we present the main proof for Lemma 1. More details can be referred to [26].

Proof (1) First, prove the following equation:

$$\hat{\beta}_{w(i)} = \hat{\beta}_w - \frac{w_i \left(H^T W H \right)^{-1} h_i e_i}{1 - w_i q_{w,i}} \tag{21}$$

where $\hat{\beta}_{w(i)}$ presents the output weights without the i observation.

For the W-ELM algorithm, the output weights are calculated as

$$\hat{\beta}_{w(i)} = \left[H_{(i)}^T W_{(i)} H_{(i)} \right]^{-1} H_{(i)} W_{(i)} T_{(i)} \tag{22}$$

Considering *A* is a nonsingular matrix with the dimension of $n \times n$, *u* and *v* are $n \times 1$ vectors, one can get

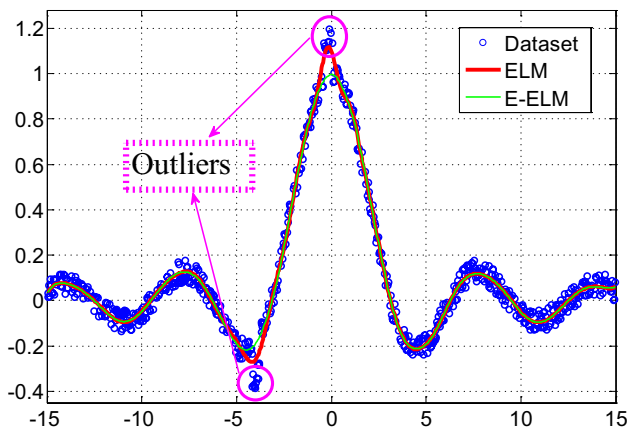$$\left(A - uv^T \right)^{-1} = A^{-1} + \frac{A^{-1} uv^T A^{-1}}{1 - u^T A^{-1} v} \tag{23}$$

Fig. 6 Comparison results of 'Sinc' function

Then

$$\left[H_{(i)}^T W_{(i)} H_{(i)}\right]^{-1} = \left[H^T WH - h_i w_i h_i^T\right]^{-1} = \left(H^T WH\right)^{-1}$$
$$+ \frac{h_i w_i h_i^T \left(H^T WH\right)^{-1}}{1 - w_i q_{w,i}}$$

Combining (22) and (23), we can obtain (21).

(2) If the ith observation is an outlier, it would present the characteristic of non-random draft. Then the model can be modified as

$$T = H\beta_w + d_i\eta + \varepsilon \tag{24}$$

where d_i is a zero vector except that the ith element is 1. It is easy to get the least squares estimation of β_w and η as

$$\hat{\beta}_w = \hat{\beta}_{w(i)}, \quad \hat{\eta} = \frac{1}{1 - w_{i,q_{w,i}}}$$

The unconstrained residual sum of squares (RSS) of model (24) can be calculated as

$$RSS = T^T T - H\hat{\beta}_w T - \hat{\eta} d_i^T T = (N - \tilde{N})\hat{\sigma}^2 - \frac{w_i e_i^2}{1 - w_{i,q_{w,i}}}$$

Then one can get

$$\hat{\sigma}_{(i)}^2 = \frac{(N - \tilde{N})\hat{\sigma}^2 - w_i e_i^2 / (1 - w_i q_{w,i})}{N - \tilde{N} - 1}$$

Part B: simulation results using well-known databases

In this part, we verify the performance of the proposed E-ELM and EW-ELM algorithms using the well-known databases. First, a function approximation problem based on the 'Sinc' function is applied to make comparison between ELM and E-ELM algorithms. In the experiment, 2000 data with 5% white noise are created. For the outliers, we make the values of the observations around $x = -4$ minus 0.15, while the data around $x = 0$ are added by 0.15. We chose 1300 observations as the training set artificially, and the last 700 data is employed for testing.

Here the number of hidden nodes is set as 50, and the active function is the sigmoid function. Figure 6 presents the simulation result, where red line represents the ELM result, while the green line is the E-ELM output. We ring out outliers with pink circles. It is obvious that E-ELM can well get rid of the interference of outliers, and obtain more accurate performance than the ordinary ELM algorithm.

Next, we employ the proposed ELM frameworks into two real-world applications: one regression application for Abalone dataset and one classification dataset for Breast Cancer dataset. All the datasets are obtained from the well-known UCI machine learning repository [33]. The OR-ELM algorithm proposed in [30] is compared with E-ELM and EW-ELM algorithms. Similar to Ref. [30], we make segmental training observations to be contaminated in order to generate outliers. Table 4 presents the simulation results of these two applications, where one can see that the proposed ELM algorithms takes a relatively longer training time than OR-ELM algorithm. However, it obtains more satisfied testing performance with the increase of contamination rate to the training observations. For the industrial applications with big delays like the silicon content

Table 4 The performance of E-ELM and EW-ELM algorithm in real world applications

Applications	Algorithms	#Nodes	Training time	Testing accuracy (RMSE) Contamination rate (%)		
				0	10	40
Abalone	ELM	200	0.0068	2.1804	2.2510	2.4502
	OR-ELM		0.1103	2.1910	2.2015	2.2513
	E-ELM		0.1732	2.1899	**2.1900**	**2.2002**
Breast Cancer	ELM	400	0.0048	96.12%	94.34%	90.78%
	OR-ELM		0.0985	95.99%	95.65%	95.46%
	EW-ELM		0.1789	95.67%	**95.68%**	**95.59%**

The numbers with bold font means the best simulation results

prediction during iron-making process in this paper, the model accuracy is more important than calculating time. So the proposed ELM frameworks are more suitable in the problems of the paper.

References

1. Gao CH, Ge QH and Jian L (2014) Rule extraction from fuzzy-based blast furnace SVM multiclassifier for decision-making. IEEE Trans Fuzzy Syst 22:586–596
2. Chen XZ, Wei JD, Xu D (2012) 3-dimension imaging system of burden surface with 6-radars array in a blast furnace. ISIJ Int 52:2048–2054
3. Bhattacharaya T (2005) Prediction of silicon content in blast furnace hot metal using partial least squares (PLS). ISIJ Int 45:1943–1945
4. Jiao KX, Zhang JL, Liu ZJ, Liu F, Liang LS (2016) Formation mechanism of the graphite-rich protective layer in blast furnace hearths. Int J Miner Metall Mater 23:16–24
5. SAXéN H, Pettersson F (2007) Nonlinear prediction of the hot metal silicon content in the blast furnace. ISIJ Int 47:1732–1737
6. Jian L, Gao CH (2013) Binary coding SVMs for the multiclass problem of blast furnace system. IEEE Trans Ind Electron 60:3846–3856
7. Nick RS, Tillander A, Jonsson TL (2013) Mathematical model of solid flow behavior in a real dimension blast furnace. ISIJ Int 53:979–987
8. stermark R, Saxén H (1996) VARMAX-modelling of blast furnace process variables. Eur J Oper Res 90: 85–101
9. Cao JW, Zhang K, Luo M, Yin C, Lai X (2016) Extreme learning machine and adaptive sparse representation for image classification. Neural Netw 81:91–102
10. Zeng JS, Gao CH (2009) Improvement of identification of blast furnace ironmaking process by outlier detection and missing value imputation. J Proc Control 19:1519–1528
11. Luo SH, Liu XG, Zeng JS (2007), Identification of multi-fractal characteristics of silicon content in blast furnace hot metal. ISIJ Int 47:1102–1107
12. Waller M, Saxén H (2000) On the development of predictive models with applications to a metallurgical process. Ind Eng Chem Res 39:982–988
13. Jian L, Gao CH, Xia ZQ(2011) A sliding-window smooth support vector regression model for nonlinear blast furnace system. Steel Res Int 82:169–179
14. Tang X, Zhuang L, Jiang C (2009) Prediction of silicon content in hot metal using support vector regression based on chaos particle swarm optimization. Expert Syst Appl 36:11853–11857
15. Zeng JS, Liu XG, Gao CH, Luo SH, Jian L (2008) Wiener model identification of blast furnace ironmaking process. ISIJ Int 48:1734–1738
16. Zheng DL, Liang RX, Zhou Y, Wang Y (2003) A chaos genetic algorithm for optimizing an artificial neural network of predicting silicon content in hot metal, J Univ Sci Technol Beijing 10:68
17. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489–501
18. Huang GB, Zhou HM, Ding XJ, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern Part B Cybern 2:513–529
19. Mao W, Wang J, Xue Z (2016) An ELM-based model with sparse-weighting strategy for sequential data imbalance problem. Int J Mach Learn Cyber. doi:10.1007/s13042-016-0509-z
20. Zong WW, Huang GB, Chen YQ (2012) Weighted extreme learning machine for imbalance learning. Neurocomputing 101:229–242
21. Zhang HG, Yin YX, Zhang S (2016) An improved ELM algorithm for the measurement of hot metal temperature in blast furnace. Neurocomputing 174:232–237.
22. Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. Int J Mach Learn Cyber 2:107–122
23. Zhou XR (2014) Online Regularized and Kernelized Extreme Learning Machines with Forgetting Mechanism, Mathematical Problems in Engineering, Article ID 938548.
24. Pearson RK (2002) Outliers in process modeling and identification. IEEE Trans Control Syst Technol 10:55–63
25. Fletcher R (1981) Practical methods of optimization: vol. 2 constrained optimization. New York
26. Sohn BY, Kim GB (1997) Detection of outliers in weighted least squares regression. Korean J Comp Appl Math 4:441–452
27. Sohn BY (1994), Weighted least squares regression diagnostics and its application to robust regression, Doctoral Thesis. Dept. of Statistics Korea University, Seoul
28. Limo K (1996) Robust error measure for supervised neural network learning with outliers. IEEE Trans Neural Netw 7:247–250
29. Jian L, Gao CH, Li L, Zend JS (2008), Application of least squares support vector machines to predict the silicon content in blast furnace hot metal. ISIJ Int 48:1659–1660
30. Zhang K, Luo MX (2015) Outlier-robust extreme learning machine for regression problems. Neurocomputing 151:1519–1527
31. Ankur M, Bikash CP (2016) Bad data detection in the context of leverage point attacks in modern power networks. IEEE Trans Smart Grid. doi:10.1109/TSG.2016.2605923
32. Cao J, Wang W, Wang J, Wang R (2016), Excavation equipment recognition based on novel acoustic statistical features. IEEE Trans Cybern. doi:10.1109/TCYB.2016.2609999
33. Bache K, Lichman M (2013) UCI Machine Learning Repository, (http://archive.ics.uci. edu/ml), School of Information and Computer Sciences, University of California, Irvine